Modelling for Science and Engineering
Data Visualitzation and Modelling

BAYESIAN NETWORKS

Excercises Block 3

1. Consider the example of the "Noisy OR-Gate model" of the Figure 1:

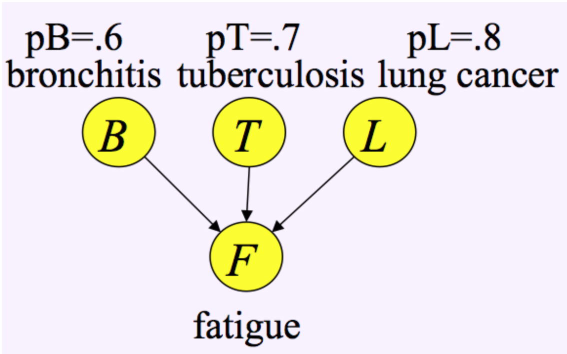


Figura 1: Causal strengths for the fatigue example.

Obtain all the conditional probabilities of node F with respect to its parents.

2. Consider the example of the "Leaky Noisy OR-Gate model" of the Figure 2:

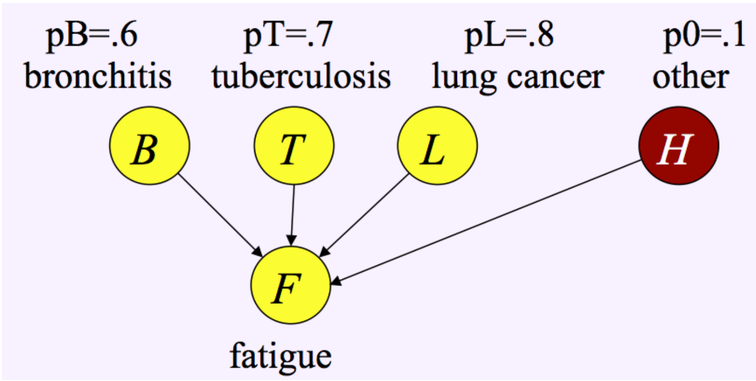


Figura 2: Causal strengths for the fatigue example with a hidden cause.

Obtain all the conditional probabilities of node F with respect to its parents in this model, and compare with the conditional probabilities obtained in Exercise 1.

3. We are interested in estimate the relative frequency of smokers on the sun terrace of a particular bar, say θ. We have decided to log whether or not individuals smoke. Of the 10 individuals there on the terrace, we obtain:

$$\{1, 2, 2, 2, 2, 1, 2, 2, 2, 1\}$$

where 1 means the individual smokes and 2 means the individual does not smoke.

Obtain the ML estimation of θ from this information.

4. Suppose we are going to sample 100 individuals who have smoked two packs of cigarettes or more daily for the past 10 years. We determine whether each individual's systolic blood pressure (X) is ≤ 100, $101-120$, $121-140$, $141-160$ or ≥ 161, and obtain the following results:

Blood Preasure Range	# of Individuals in this Range
≤ 100	2
$1001-120$	15
$121-140$	23
$141-160$	25
≥ 161	35

Let denote

$$\begin{aligned}\theta_1 &= P(X \leq 100)\\ \theta_2 &= P(101 \leq X \leq 120)\\ \theta_3 &= P(121 \leq X \leq 140)\\ \theta_4 &= P(141 \leq X \leq 160)\\ \theta_5 &= P(X \geq 161)\end{aligned}$$

Obtain the ML estimations of θ_i, $i = 1, \ldots, 5$, from this information.

5. Consider the following Bayesian Networks for parameter learning:

Bayesian Network A: $V = \{X, Y\}$, $E = \{(X, Y)\}$, and

Bayesian Network B: $V = \{X, Y\}$, $E = \{(Y, X)\}$,

where both X and Y are boolean random variables. X takes values x_1 and x_2, and variable Y takes values y_1 and y_2.

(a) For each of the BN, specify which are the parameters.

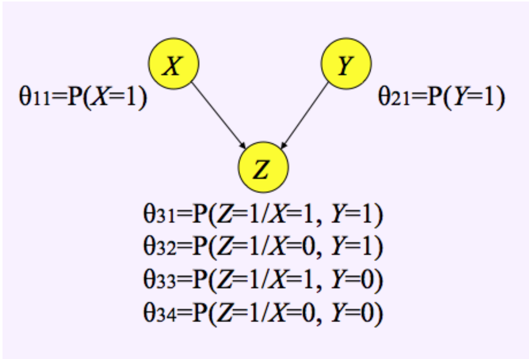
(b) For each of the BN, compute the MLE for the parameters specified in (a) from the following data set D.

Case	X	Y
1	x1	y1
2	x1	y1
3	x1	y1
4	x1	y1
5	x1	y1
6	x1	y2
7	x2	y1
8	x2	y1
9	x2	y2
10	x2	y2

(c) For each of the BN estimate $P(X = x_1)$ and $P(Y = y_1)$ from the MLE estimations obtained in (b), and compare their values for the two BNs.

(d) If we have the evidence that $X = x_1$, which is the prediction for variable Y, and with which confidence level, by using each of the two Bayesian Networks?

6. Consider the following Bayesian network, where the three variables are binary (take values 1 and 0):



(a) From complete data D given by the following table, compute the MLE of the parameters $\theta = (\theta_{11},\, \theta_{21}, \theta_{31},\, \theta_{32},\, \theta_{33},\, \theta_{34})$.

Case	X	Y	Z
1	1	1	1
2	1	1	0
3	1	1	0
4	1	0	1
5	0	1	1
6	0	1	0
7	0	0	1
8	0	0	1
9	0	0	0
10	0	0	0

Also estimate the probability $P(Z = 1)$.

(b) Imagine now that the available data in the previous section were incomplete, as shown in the following table, corresponding to data ID_1.

Case	X	Y	Z
1	?	1	1
2	1	1	0
3	1	1	0
4	1	0	?
5	0	1	1
6	0	1	0
7	0	0	1
8	0	?	1
9	0	0	0
10	0	0	0

Therefore, we want to apply the Expectation Maximization (EM) method to estimate the parameters. If we start with the initial estimates of the parameters

$$\theta^0 = (\theta^0_{11} = 0.3,\, \theta^0_{21} = 0.6,\, \theta^0_{31} = 0.4,\, \theta^0_{32} = 0.6,\, \theta^0_{33} = 0.8,\, \theta^0_{34} = 0.7),$$

i. Compute the likelihood function for this initial estimates of the parameters from the incomplete data we have, ID_1.

ii. Obtain the expected empirical distribution for the variables X, Y and Z corresponding to the completations of the incomplete data ID_1 and the corresponding probabilities based on the initial estimates of the parameters θ^0.

iii. Use the expected empirical distribution found in the previous item to find the estimates of the parameters corresponding to iteration 1, $\hat{\theta}^1$.

iv. Compute the likelihood function for the first iteration estimates of the parameters $\hat{\theta}^1$ from the incomplete data ID_1, and compare it with that obtained in item i).

(c) Repeat item (b) but now with the incomplete data ID_2 in the following table, instead of ID_1.

Case	X	Y	Z
1	?	?	1
2	1	1	0
3	1	1	0
4	1	0	1
5	0	1	1
6	0	1	0
7	0	0	1
8	0	0	1
9	0	0	0
10	0	0	0

7. (**Exercise to deliver individually. Please, attach R scripts**)

Load data file "datos_generados_ejercicio_7.rdata" with R. Then, you will have a data frame with 500 cases corresponding to 3 variables: X, Y and Z. Learn the parameters of the graph structure given in the picture of Exercise 6, by using bnlearn package, from data.

Remark: bnlearn can no deal with "NA". So, first of all, omit cases with 'NA"(then, from the initial 500 cases you will finally have 420).